

Claims:

1. A method of processing a data stream, comprising:
 - receiving a set of elements of said data stream;
 - storing a data structure in a memory, said data structure configured to represent said set of elements as a hierarchy of nodes, each of said nodes having frequency data associated with one of: an element in said set of elements or a prefix of an element in said set of elements; and
 - processing said data structure to identify a set of hierarchical heavy hitter nodes among said nodes, said frequency data of each of said hierarchical heavy hitter nodes, after discounting any portion thereof attributed to a descendant hierarchical heavy hitter node in said set of hierarchical heavy hitter nodes, being greater than or equal to a fraction of the number of elements in said set of elements.
2. The method of claim 1, wherein said data structure comprises a trie data structure.
3. The method of claim 2, wherein said step of storing comprises, for each element in said set of elements, at least one of:
 - creating at least one node in said hierarchy of nodes; and
 - incrementing said frequency data of at least one node in said hierarchy of nodes.
4. The method of claim 3, wherein said step of storing further comprises:
 - compressing said trie data structure by deleting one or more nodes in said hierarchy of nodes where said frequency data thereof is less than a predefined threshold.
5. The method of claim 1, wherein said data structure comprises a sketch-based summary structure.
6. The method of claim 5, wherein said step of storing comprises:

creating a plurality of subsets, each of said plurality of subsets being associated with one or more elements in said set of elements; and
associating a counter with each of said plurality of subsets.

7. The method of claim 6, wherein said step of storing further comprises, for each element in said set of elements, one of:

incrementing said counter associated with each subset of said plurality of subsets having said element; and

decrementing said counter associated with each subset of said plurality of subsets having said element.

8. The method of claim 6, wherein for each subset of said plurality of subsets, a probability that any of said elements in said set of elements is in said subset is a fixed value.

9. Apparatus for processing a data stream, comprising:

means for receiving a set of elements of said data stream;

means for storing a data structure configured to represent said set of elements in a memory as a hierarchy of nodes, each of said nodes having frequency data associated with one of: an element in said set of elements or a prefix of an element in said set of elements; and

means for processing said data structure to identify a set of hierarchical heavy hitter nodes among said nodes, said frequency data of each of said hierarchical heavy hitter nodes, after discounting any portion thereof attributed to a descendant hierarchical heavy hitter node in said set of hierarchical heavy hitter nodes, being greater than or equal to a fraction of the number of elements in said set of elements.

10. The apparatus of claim 9, wherein said data structure is a trie data structure.

11. The apparatus of claim 10, wherein said means for storing comprises:

means for creating at least one node in said hierarchy of nodes; and

means for incrementing said frequency data of at least one node in said hierarchy of nodes.

12. The apparatus of claim 11, wherein said means for storing further comprises:

means for compressing said trie data structure by deleting one or more nodes in said hierarchy of nodes where said frequency data thereof is less than a predefined threshold.

13. The apparatus of claim 9, wherein said data structure is a sketch-based summary structure.

14. The apparatus of claim 13, wherein said means for storing comprises:

means for creating a plurality of subsets, each of said plurality of subsets being associated with one or more elements in said set of elements; and

means for associating a counter with each of said plurality of subsets.

15. A computer readable medium having stored thereon instructions that, when executed by a processor, cause the processor to perform a method of processing a data stream, comprising:

receiving a set of elements of said data stream;

storing a data structure in a memory, said data structure configured to represent said set of elements in a memory as a hierarchy of nodes, each of said nodes having frequency data associated with one of: an element in said set of elements or a prefix of an element in said set of elements; and

processing said data structure to identify a set of hierarchical heavy hitter nodes among said nodes, said frequency data of each of said hierarchical heavy hitter nodes, after discounting any portion thereof attributed to a descendant hierarchical heavy hitter node in said set of hierarchical heavy hitter nodes, being greater than or equal to a fraction of the number of elements in said set of elements.

16. The computer readable medium of claim 15, wherein said data structure is a trie data structure.

17. The computer readable medium of claim 16, wherein said step of storing comprises, for each element in said set of elements, at least one of:

creating at least one node in said hierarchy of nodes; and
incrementing said frequency data of at least one node in said hierarchy of nodes.

18. The computer readable medium of claim 17, wherein said step of storing further comprises:

compressing said trie data structure by deleting one or more nodes in said hierarchy of nodes where said frequency data thereof is less than a predefined threshold.

19. The computer readable medium of claim 15, wherein said data structure is a sketch-based summary structure.

20. The computer readable medium of claim 19, wherein said step of storing comprises:

creating a plurality of subsets, each of said plurality of subsets being associated with one or more elements in said set of elements; and
associating a counter with each of said plurality of subsets.

21. The computer readable medium of claim 20, wherein said step of storing further comprises, for each element in said set of elements, one of:

incrementing said counter associated with each subset of said plurality of subsets having said element; and
decrementing said counter associated with each subset of said plurality of subsets having said element.

22. The computer readable medium of claim 20, wherein for each subset of said plurality of subsets, a probability that any of said elements in said set of elements is in said subset is a fixed value.